European Commission

JRC TECHNICAL REPORTS

A Social Multi-Criteria Framework for *ex-ante* Impact Assessment

Operational Issues

Munda, Giuseppe

2017

Joint Research Centre

EUR 28752 EN

This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication.

Contact information
Giuseppe Munda
European Commission, Joint Research Centre
Directorate I – Unit JRC.I.1 – Modelling, Indicators and Impact Evaluation
TP 361 – Via E .Fermi 2749 – I-21027 – Ispra (Va) – ITALY
giuseppe.munda@ec.europa.eu

JRC Science Hub
https://ec.europa.eu/jrc

JRC107899

EUR 28752 EN

PDF ISBN 978-92-79-72293-6 ISSN 1831-9424 doi:10.2760/909528

Luxembourg: Publications Office of the European Union, 2017

How to cite this report: Munda G. *A social multi-criteria framework for ex-ante impact assessment,* EUR 28752 EN, Publications Office of the European Union, Luxembourg, 2017, ISBN 978-92-79-72293-6, doi:10.2760/909528, JRC107899

Contents

Acknowledgements

Comments by Leen Hordijk, Sven Langedijk, Marion Le Louarn and Paul Smits are gratefully acknowledged. The usual disclaimer applies.

Note

This report has been developed in the context of the activities of the Competence Centre on Modelling.

Abstract

European Commission's current practice of Impact Assessment (IA) considers three main objectives i.e. efficiency, effectiveness (including proportionality) and coherence and it is based on the assessment of various broad impacts such as economic, environmental and social (including distribution of costs and benefits among social actors) ones. There is no doubt that IA is multidimensional in nature and as a consequence, multi-criteria evaluation (MCE), and in particular social multi-criteria evaluation (SMCE), which has been explicitly designed for public policy, can be a very useful methodological and operational framework. SMCE tries to integrate different scientific approaches with social actors' preferences, thus being a consistent and transparent framework for both inter-disciplinarity and public participation. This report aims at presenting:

- A *methodological framework* where the hierarchical structure of the option comparison step of a typical *ex-ante* IA (including dimensions, objectives and evaluation criteria) is clarified as much as possible by means of well-established concepts in the decision theory literature. This might help in increasing the degree of homogeneity across IA studies.

- A *measurement framework* where

1. the various criterion scores can assess impacts by using both quantitative (e.g. as result of simulation models) and qualitative (e.g. results of participatory techniques) information, and
2. the mathematical aggregation rule guarantees consistency and transparency of results.

- An illustrative example dealing with a recent IA on modernising VAT for cross-border B2C e-Commerce.

1 Introduction

European Commission's current practice of Impact Assessment (IA) considers three main objectives i.e. efficiency, effectiveness (including proportionality) and coherence. Moreover IA includes various broad impacts such as economic, environmental and social ones. There is no doubt that IA is multidimensional in nature, and as a consequence, multi-criteria evaluation (MCE), and in particular social multi-criteria evaluation (SMCE), which has been explicitly designed for public policy, can be a very useful methodological and operational framework. The main achievement of SMCE is the fact that the use of various evaluation criteria has a direct translation in terms of plurality of values and dimensions used in the evaluation exercise. SMCE accomplishes the goals of being *inter/multi-disciplinary* (with respect to the research team), *participatory* (with respect to the community) and *transparent* (since all criteria are presented in their original form without any transformations in money, energy or whatever common measurement rod) (Munda, 2004, 2008, 2016).

Of course, policy *ex-ante* evaluation/assessment is not a one-shot activity. On the contrary, it takes place as a *learning process* which is usually highly dynamic, so that judgements regarding the political relevance of items, alternatives or impacts may present sudden changes, hence requiring a policy analysis to be flexible and adaptive in nature. This is the reason why evaluation/assessment processes have a *cyclic nature*. By this is meant the possible adaptation of elements of the process due to continuous feedback loops among the various steps and consultations among the actors involved. In this framework, mathematical models still play a very important role, i.e. the one of guaranteeing consistency between assumptions used and results obtained. In general, in a multi-criterion problem, there is no solution optimising all the criteria at the same time (*ideal or utopia solution*), and therefore *"compromise solutions"* have to be found (Roy, 1996).

Let's introduce multi-criteria evaluation by means of a simple everyday life example. Let's imagine staying in front of a shop and looking at a set of jackets one likes. Which will be the next step? Probably enter the shop and ask for the *price*. At this point, we have two possibilities: to leave the shop because we think that the price is too high or to accept the price as a reasonable one. In the second case, we still need to choose the jacket we want from the original set (of e.g. ten jackets). Thus, probably we are going to try the jackets and see which one *fits better aesthetically to our body.* Let's assume that we are still undecided among four of them and for sure we do not like the other six. How do we choose among the four which are indifferent? Maybe at this stage we will use the criterion *colour.* Let's imagine we are still indifferent between two jackets. Probably now we will look at the *quality of the textile composition* and we will finally choose the one with the higher quality. This is an example of a selection of a final alternative by using a peculiar multi-criterion aggregation rule, called the *lexicographic model*. This model refers to the procedure used to put in order the words in a dictionary, the first letter playing the role of the first criterion, the second letter, the second criterion, and so on. In the lexicographic model, all actions are first ranked by means of the first criterion, and then if some indifferent actions exist, these are further explored by means of the second criterion, and so on. Lexicographic orders usually lead to a straightforward selection of the most preferred alternative; however, most of the information collected on alternatives will not play a role in the evaluation process.

Let's discuss this example a bit and draw some conclusions. First of all, did we experience sometimes a decision process like this one? Probably yes, even if not with these criteria or in this order. Thus apparently human beings use multi-criteria evaluation without any formal knowledge of it; we could then state that it is *a behavioural assumption with a high degree of real-world descriptive content.* Secondly, does the order of criteria have any influence on the final alternative selected? Of course it does. If one starts with the quality criterion instead of price the jacket selected will probably be the most expensive one. This shows that when using various criteria human beings do not necessarily attach to them the same *weight*. In the case of the lexicographic model, in principle the first criterion alone could be enough to select the preferred solution (i.e. if only one jacket has the price we are willing to pay), this implies that its weight is much higher than any other criterion used in the selection process. This is the reason why the first criterion is sometimes called the "dictator". Clearly then, the order of consideration of criteria determines their relative weights. Thirdly, what happens in our example if we do not like the overall characteristics of the selected jacket finally? Probably we will start again the process e.g. changing the order of criteria (i.e. their weight), thus for example, accepting to pay a higher price. Again this is something that we have probably experienced and which shows that what really matters is the *learning process* and not the alternative finally selected. This latter one is *constructed* by means of the decision process and not discovered as a global optimum.

Finally, does the lexicographic method allow for any *compensability* among the various criteria considered? Intuitively, compensability refers to the possibility that some bad criterion scores can be compensated by other very good criterion scores. For example, an overall student evaluation can be based on the principle that a very bad score in mathematics (let's say a 2 in a 0-10 scale) can be compensated by a 10 in literature and thus the student can pass the final evaluation. This evaluation system is a *completely compensatory* one. On the contrary, an evaluation system can alternatively be based on the principle that a student has to be "enough good" in all the subjects and a 2 in mathematics cannot be compensated by any other score, however high. This second evaluation system would be a *partial compensatory* one. Compensability then assumes a certain degree of mutual interaction among the criterion scores, if no such interaction is possible, no compensability exists. Since in a lexicographic method the evaluation criteria are not considered simultaneously, this procedure is completely *non-compensatory*. Compensability is a very important concept when multi-criteria evaluation is applied to integrate various policy dimensions. For example, in evaluating a policy option, if we consider that 2 in mathematics could be a very bad environmental impact and 10 in literature a very good economic impact, it is clear that allowing or not for compensability and to which degree is the central issue in evaluation exercises.

To search for compromises implies that no-dictator must exist. That is, all the criteria relevant in a policy problem have to be used simultaneously and not in a lexicographic order, since otherwise some criteria will have a much higher weight a priori. Thus for example, a legislative system which foresees that a financial analysis of projects has to be done before the evaluation of their environmental impacts, it is indeed prioritizing the economic dimension with respect to the environmental one. *Multi-criteria evaluation for ex-ante evaluation/impact assessment must then be based on more general models than the lexicographic one, allowing the use of different objectives and criteria at the same time*.

The present report aims at developing a SMCE operational framework useful for answering the following key question: *how can the Commission integrate a plurality of technical aspects and social views into its ex-ante impact assessment in a coherent and transparent manner (coherence and transparency being key Better Regulation requirements)?* Section 2 illustrates the main concepts of MCE and their adaptation to structure an IA exercise. Section 3 presents a measurement framework where the various criterion scores can be both qualitative and quantitative, and the mathematical aggregation rule is as consistent and simple as possible. Section 4 presents an illustrative example and finally some conclusions are drawn.

2 The Hierarchical Structure of *ex-ante* Impact Assessment

Commission Impact Assessment is based on three main objectives:

1. effectiveness (i.e. the degree to which the policy objectives are achieved (in terms of goals or levels of output) and the problems identified are solved,
2. efficiency (the EC has expressed its commitment to ensure that its proposals meet policy goals at minimum cost and/or taking into account an analysis of costs and benefits, and their distribution among the stakeholders affected) and
3. coherence (with other existing EU policies).

Correctly both the objectives of effectiveness and of efficiency are considered jointly; otherwise there is the risk to drive the policy evaluation framework towards a situation in which efficiency would be privileged at the cost of effectiveness (focusing on just "cheap" options, regardless of the level of output/outcomes achieved).

The objective of fairness is not considered explicitly, although probably it should be dealt with too. Given the existence of a plurality of social actors, who have different stakes in the policy being assessed, a conflictual situation frequently arises; thus distributional issues always play a central role - hence the BR guidelines/toobox's recommendation of not assessing only global efficiency but also to ensure that the distribution of costs and benefits among the stakeholders affected is deemed acceptable.

Any social decision problem is characterized by conflicts between competing values, perspectives, interests and different groups and communities that represent them. Any policy option generally implies winners and losers, thus it is important to check if a policy option might be preferred just because some dimensions (e.g. the environmental) or some social groups (e.g. the lower income groups) are not taken into account in the analysis; ignoring some evidently existing dimensions is considered more and more unacceptable in a public policy context in general, and in the Commission in particularly - as e.g. recently emphasized in the European Pillar of Social Rights. In my opinion, explicit inclusion of fairness within an IA framework mainly implies that a) social values, interests and desires should be considered as much as possible, b) distributional issues should be explored at the highest degree possible and c) the whole ex ante evaluation/impact assessment process should be transparent.

Let us now look at the options' comparison step in Commission IAs, such as the ones represented in Tables from 1 to 8, taken from published IA studies.

	Overall impact			Key objectives		
Policy option	Social	Economic	Environmental	Effectivness	Efficiency	Coherence
Mainstreaming renewables in the heating and cooling supply						
Option 0-partial continuation of current RED requirement + EPBD + EED	0	0	0	0	0	0
Option 1-RES H&C obligation for fossil fuel suppliers	--	+	++	++	+	+
Option 2-RES H&C obligation on all fuel suppliers	-	+	++	++	++	+
Facilitating the uptake of renewable energy and waste heat in DHC systems						
Option 0-baseline	0	0	0	0	0	0
Option 1-continuation of current requirements, with best practice sharing	+	0	0	-	0	0

Table 1. Example of comparison of options to increase the use of renewable energy

Source: SWD(2016) 418 final, p. 119

Criteria	**Baseline**	**Option 2** **1 ppm**	**Option 3** **0.5 ppm**	**Option 4** **5 ppm**
Effectiveness	0	≈/+	+	≈
Efficiency	0	-	--	≈
Coherence	0	+	+	≈
Scientific advice (SCOEL)	Additional leukaemia risk with 1ppm exposure for a 40-year working life: from 0 to 10.78 extra leukaemia deaths between ages 25-85 years.			
ACSH	1 ppm (=2.25 mg/m3) to be reviewed. A revision to be conducted in 3 years.			

Table 2. Example of comparison of options to protect workers from leukaemia risk

Source: SWD(2016) 152 final, p. 41

Alternatives	**Impacts**		
	Economic	**Environmental**	**Social**
Scenario 1	3-77 million EUR/y	++	0
Scenario 2	13-135 million EUR/y	+++	+
Scenario 3	71-135 million EUR/y	+++	-

Table 3. Example of comparison of options for mercury regulation

Source: SWD(2016) 17 final, p. 44

Operational objectives	Baseline B	Baseline B bis	Simple	Limited changes	Targeted
Technological progress reflected	0	0	-	0	++
Incentives to innovate fully preserved	0	0	+	0	-
No undue costs for most efficient installations	0	0	-	++	++
Better alignment with production levels	0	0	+	++	++
Avoid windfall profits	0	0	-	+	++
No increased administrative complexity	0	0	++	-	--

Table 4. Example of comparison of options for greenhouse gas emission reduction

Source: SWD(2015) 135 final, p. 52

	Baseline (directive unchanged / no support for innovation)	**Alternative baseline** (Current rules continued)	**Option 1** (Amended approach for all sectors with tailoring for industry)	**Option 2** (Permanent financing facility)
Minimise complexity and administrative burden	Not applicable	0	-	-
Complementarity with other EU instruments	Not applicable	0	+	+
Effectiveness in addressing barriers for low-carbon innovation	Not applicable	0	++	+
Potential to attract innovative projects	Not applicable	0	++	0
Leverage	Not applicable	0	-	+
Competitiveness	Not applicable	0	++	++
EU added value and geographical distribution	Not applicable	0	+	+

Table 5. Example of comparison of options for the Innovation Fund

Source: SWD(2015) 135 final, p. 69

	Increase effectiveness	Increase coherence	Minimise risk of market distortion	Minimise administrative burden
Baseline	No Modernisation Fund	No Modernisation Fund	No Modernisation Fund	No Modernisation Fund
Option 1	+	0	-	-
Option 2	++	+	+	+
Option 3	0	+	++	+

Table 6. Example of comparison of options for the Modernisation Fund

Source: SWD(2015) 135 final, p. 81

	Option 1	Option 2
Impact on GDP, NPV 2015-2055 (change compared to baseline)	+ € 8.3 billion (+0.002%)	+ € 13.3 billion (+0.003%)
Impact on labour force in 2030/2050 (change compared to baseline)	8 000/10 000 (0.00%)	11 000/13 000 (0.00/0.01%)
Impact on employment in 2030/2050 (change compared to baseline)	4 000/6 000 (0.00%)	5 000/8 000 (0.00%)
Impact on real incomes in 2030/2050 (change compared to baseline)	+ € 0.2 billion /+ € 1.1 billion (0.00%)	+ € 0.6 billion /+ € 1.5 billion (0.00/0.01%)

Table 7. Example of macro-economic comparison of options for maternity leave

Source: SWD(2017) 202 final, p. 77

Options	Effectiveness	Efficiency	Coherence
Baseline	0	0	0
Non-legislative Option	+ Increased retention of women in the labour market	+++ Very limited costs and positive impacts on companies and Member States	+ Positive impact on fundamental rights of pregnant women.
Option 1	++ Strong increase in the retention of women in the labour market	++ Positive impacts on companies and Member States	++ High positive impact on fundamental rights of pregnant women.
Option 2	++ Strong increase in the retention of women in the labour market	+/- Positive impacts on companies and negative on Member States	++ Highest positive impact on fundamental rights of pregnant women.

Table 8. Example of comparison of options for maternity leave

Source: SWD(2017) 202 final, p. 78

As one can see there is some inconsistency in the terminology used. For example, efficiency, effectiveness and coherence are classified as objectives or criteria across different studies. The same applies to social, economic and environmental impacts, which may be classified as criteria or objectives, while other studies use more specific evaluation criteria. Let us then introduce some standard concepts from decision theory literature, which can help in eliminating confusion and in increasing homogeneity across IA studies (see e.g. Figueira *et al.,* 2016; Roy, 1996; Vincke, 1992).

"Dimension" is the highest hierarchical level of analysis and indicates the scope of objectives, criteria and criterion scores. In IA studies, the general categories of economic, social and environmental impacts should be considered as dimensions.

"Objective" indicates the direction of change desired. For example, within the economic dimension economic growth has to be incentivised; within the environmental dimension contribution to the EU's climate change commitments in the context of COP 21 has to be maximised; in the energy dimension, energy security should be maximised. In impact assessment, effectiveness, efficiency and coherence have to be considered objectives clearly. In the example shown in table 5, "minimise complexity .." is an objective and not an evaluation criterion.

"Evaluation criterion" is the basis for evaluation in relation to a given objective (any objective may imply a number of different criteria). It is a function that associates each policy option with a variable indicating its desirability according to expected consequences related to the same objective, a classical example in the economic dimension might be GDP, saving rate and inflation rate inside the objective "growth maximization". The examples shown in Tables 4 and 7 are based on this notion of evaluation criteria.

"Criterion score" is an assessment of the impact consistent with a given criterion with reference to a policy option to be evaluated. It is a measurement deriving from a process that represents, at a given point in space and time, a shared perception of a real-world state of affairs. To give an example, when comparing two countries, within the economic dimension, one objective could be "maximization of economic growth"; the criterion might be research and development performance, the criterion score could be "number of patents per million of inhabitants". Criterion scores can be both qualitative (as in most of current IA studies) or quantitative. Uncertainty (stochastic or fuzzy) can also be included.

"Constraint" is a limit on the values that criterion scores may assume; it may or may not be stated mathematically. For example, in IA, political feasibility in terms of respect of the principles of proportionality and subsidiarity can be seen as a constraint.

"Goal" is synonymous with a target and is something that can be either achieved or missed, e.g. at least 95% of children (from 4 to compulsory school age) should participate in early childhood education, the rate of early leavers from education and training aged 18-24 should be below 10%. If a goal cannot, or is unlikely to, be achieved, it may be converted to an objective. *"Attribute"* is a measure that indicates whether goals have been met or not. Attributes and goal can then be used for measuring effectiveness.

By using these concepts, a typical IA study can be represented as the hierarchical structure illustrated in Figure 1.

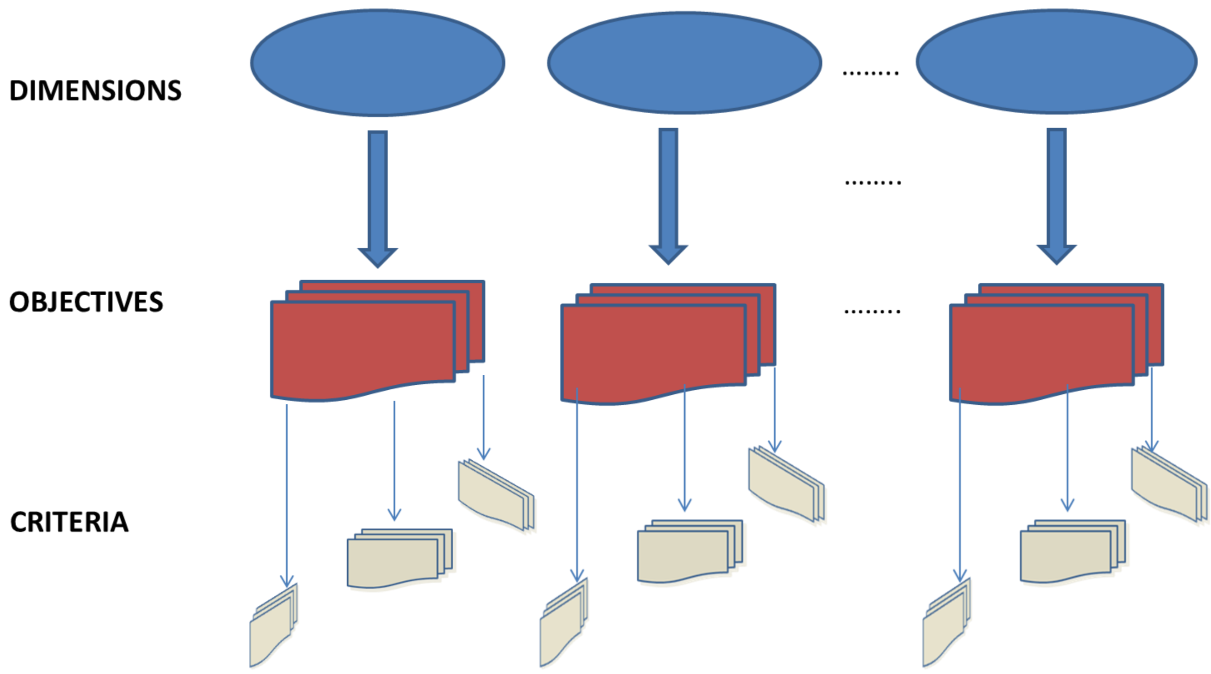

Figure 1. A Schematized Vision of the Hierarchical Structure of an *ex-ante* IA

An important consequence of this hierarchical problem structuring is on criterion weighting. A common practice is the pragmatic solution of *no criterion weighting*. This approach normally reduces conflicts in the problem structuring step, but the question here is: is it correct?

Indeed the fact that all criteria have the same weight *does not guarantee at all that objectives and dimensions have the same weight.* This would be guaranteed only under the condition that all the dimensions have the same number of criteria. This of course is quite unnatural and artificial and even dangerous. Analysts could be tempted to choose the same number of criteria for each dimension even if these criteria were completely *redundant*.

The conclusion can thus be drawn that by giving the same weight to all the criteria the different dimensions have different weights (since any dimension is then weighted according to its number of criteria). On the contrary *different criterion weights* can guarantee that all the *dimensions are considered equal*!

A reasonable practice can be to start by giving the same weight to each dimension and then splitting each weight among the objectives and criteria of any dimension proportionally[1]. Of course, one could assume that some dimensions are more important than other ones, and thus their weight should be higher, but this should be justified based on strong and transparent ethical, scientific, institutional or legal arguments.

This also implies that sensitivity or robustness analyses have to check the consequences for the final ranking of these arguments and not of all the possible combinations of weights. Sensitivity and robustness analysis are thus a way to improve transparency on the assumptions introduced in an IA study.

[1] In this framework weights are meaningful only as *importance coefficients* and not as trade-offs (see Annex 2).

3 A SMCE based Measurement Framework for *ex-ante* IA

A *"discrete multi-criterion problem"* can be formally described as follows (see e.g. Figueira *et al.,* 2016). *A* is a finite set of *N* feasible actions (or alternatives). *M* is the number of different points of view, or evaluation criteria, g_m, that are considered relevant to a specific policy problem. Where action ***a*** is evaluated to be better than action ***b*** (both belonging to the set *A*), by the *m*-th point of view, then $g_m(a)>g_m(b)$. In this way a decision problem may be represented in an *N* by *M* matrix *P* called an *evaluation or impact matrix*. In such a matrix, the typical element p_{mn} $(m=1, 2, ..., M; n=1, 2, ..., N)$ represents the evaluation of the *n*-th alternative by means of the *m*-th criterion, in other words, each criterion score represents the performance of each alternative according to each criterion (see Table 9). The impact matrix may include quantitative, qualitative or both types of information.

		Alternatives			
Criteria	**Units**	$\mathbf{a_1}$	$\mathbf{a_2}$	$\mathbf{a_3}$	$\mathbf{a_4}$
$\mathbf{g_1}$		$g_1(a_1)$	$g_1(a_2)$	.	$g_1(a_4)$
$\mathbf{g_2}$		.	.	.	.
$\mathbf{g_3}$		.	.	.	.
$\mathbf{g_4}$		.	.	.	.
$\mathbf{g_5}$		.	.	.	.
$\mathbf{g_6}$		$g_6(a_1)$	$g_6(a_2)$	.	$g_6(a_4)$

Table 9. Example of an Impact Matrix

In a discrete multi-criteria problem, there is a range of multi-criteria problem formulations, which may take one of the following forms (Roy, 1996):

(α) the aim is to identify one and only one final alternative;

(β) the aim is the assignment of each alternative to an appropriate predefined category according to what one wants it to become afterwards (for instance, acceptance, rejection or delay for additional information);

(γ) the aim is to rank all feasible alternatives according to a total or partial pre-order;

(δ) the aim is to describe relevant alternatives and their consequences.

The importance of mathematical approaches in SMCE is their ability to allow a consistent aggregation of the diverse information. Otherwise, even if everybody would agree on the considerations contained in the previous Section, their implementation in a real-world assessment exercise would be impossible. Just to give an example of the typical difficulties we may find when solving multi-criterion problems, let's look at the numerical example shown in Table 10, where 21 criteria rank four alternatives (*a, b, c, d*). Criteria are grouped according to the ranking they support (i.e. 3 criteria are in favour of *abcd*, while 7 in favour of *bdca*, and so on).

Number of criteria	**3**	**5**	**7**	**6**
	a	a	b	c
	b	c	d	b
	c	b	c	d
	d	d	a	a

Table 10 Numerical Example with 21 Criteria and 4 Alternatives

(Source: Moulin, 1988, p. 228)

We can assume that the objective is to isolate one alternative (α-decision problem formulation). A first possibility is to apply the so-called *plurality rule*, meaning that the alternative which is most often ranked in first place is the winner. Thus, in our case, alternative ***a*** would be chosen since eight criteria put it in first position. However, if one looks carefully at Table 10, it can be seen that alternative ***a*** also has the strongest opposition, since 13 criteria put it in last position.

From this plurality rule paradox two main lessons can be learned:

1. Good ranking procedures should respect the entire ranking of alternatives and not the first position only.
2. It is important to consider not only what a majority of criteria prefer, but also what they reject.

A multi-criteria mathematical procedure is an aggregate of all objectives (or goals), criteria (or attributes) and criterion scores. Given that in a multi-criterion problem, there is no ideal solution (at least in the majority of cases), assessing the *quality* of multi-criteria mathematical procedures is impossible, either based on a notion of *approximation* (i.e., discovering pre-existing truths) or on a mathematical property of *convergence* (i.e., does the decision automatically lead, in a finite number of steps, to the optimum ***a****?).

In 1986 Kenneth Arrow and Hervé Raynaud published a very influential book titled *"Social choice and multicriterion decision-making",* where the formal analogies between the multi-criterion problem and the social choice one are analysed deeply. This book is based on the assumption that, in the case where all criteria have ordinal impact scores, if one considers the evaluation criteria as voters, a multi-criteria impact matrix and a voting matrix are identical.

As a consequence all results of social choice also apply to multi-criteria evaluation fully; in particular Arrow's impossibility theorem stating that there is no perfect mathematical aggregation rule. Thus, unlike other mathematical fields, neither approximation nor convergence criteria can be used; only "reasonable" mathematical procedures can be developed in this framework. Reasonable here means that algorithms can be evaluated not only according to the *formal properties* they present, but, overall, according to the *empirical consequences* implied by their use.

Here, I will try to isolate some properties that can be considered desirable for a discrete multi-criteria aggregation rule (often called multi-criteria *method*) in the framework of Commission *ex-ante* IA. In synthesis, the information contained in the impact matrix useful for solving the so-called multi-criterion problem is:

- Intensity of preference (when quantitative criterion scores are present).
- Number of criteria in favour of a given alternative.
- Weight attached to each single criterion.
- Relationship of each single alternative with all the other alternatives.

Combinations of this information generate different aggregation conventions, i.e. manipulation rules of the available information to arrive at a preference structure. The aggregation of several criteria implies taking a position on the fundamental issue of compensability. As we have already observed in the Introduction, complete *compensability* is not desirable for the problem we are dealing with, since it implies that e.g. a good performance on efficiency would offset a very bad one on effectiveness or vice versa. On the other hand, complete non-compensability is not desirable either, since it would imply the use of a lexicographic model and the consequent choice of a "dictator", e.g. efficiency.

As a consequence, the only option left is the use of *partial compensatory* methods, such as the "outranking methods", including e.g. ELECTRE (Roy, 1996) and PROMETHEE

(Brans *et al*., 1986). These methods, following the Condorcet tradition, entail aggregating the criteria into a partial binary relation *aSb* (an outranking relation) based on concordance and discordance indexes, and then "exploiting" this relationship. Each of these two steps may be treated in a number of ways according to the problem formulation and the particular case under consideration.

To illustrate this approach consider Parliamentary voting. The concordant coalition can be considered as the sum of the votes of the members in favour of a given option; according to a majority voting rule this option will be approved if it obtains more than 50% of the votes. According to the normative tradition in political philosophy, all coalitions, however small, should be given some fraction of the decision power. One measure of this power is the ability to veto certain subsets of outcomes. This explains the use of the condition of non-discordance.

In practice, the effect of the discordance test is that even if *M-1* criteria support the recommendation of choosing ***a*** over ***b***, this recommendation must not be accepted if only one criterion is against it with a strength bigger than the veto threshold. This implies that in a situation where all criteria would support a policy option, this option cannot be accepted if one criterion is very strongly against this option. Of course, this depends on the way in which "very strongly" is defined, i.e. the definition of the veto threshold.

It has been argued that the presence of qualitative information in evaluation problems concerning socio-economic issues is a rule, rather than an exception (Nijkamp *et al*., 1990). Thus there is a clear need for methods that are able to take into account information of a "mixed" type (both qualitative and quantitative criterion scores). For simplicity, we refer to qualitative information as information measured on a nominal or ordinal scale, and to quantitative information as information measured on an interval or ratio scale.

Moreover, ideally, this information should be precise, certain, exhaustive and unequivocal. But in reality, it is often necessary to use information which does not have those characteristics so that one has to face the uncertainty of a stochastic and/or fuzzy nature present in the data. As a consequence, multi-criteria methods able to tackle consistently the widest types of *mixed information* should be considered as desirable ones.

In the 1990s some outranking methods were especially designed to address public policy analysis, one of the most widespread being NAIADE[2] (Munda, 1995). It is a discrete multi-criteria method whose impact matrix may include crisp, stochastic or fuzzy measurements of the performance of an alternative with respect to an evaluation criterion. Thus it is very flexible for real-world applications. NAIADE can give the following information:

- ranking of the alternatives according to the set of evaluation criteria (i.e. technical compromise solution/s);
- indications of the distance of the positions of the various interest groups (i.e. possibilities of convergence of interests or coalition formations);
- ranking of the alternatives according to actors' impacts or preferences (i.e. social compromise solution/s).

From a mathematical point of view, two main issues are solved:

1. the problem of equivalence of the procedures used in order to standardize the mixed criterion scores;
2. the problem of comparison of fuzzy numbers typical of all fuzzy multi-criteria methods.

[2] NAIADE (Novel Approach to Imprecise Assessment and Decision Environments), Patent Number: SN 2544 since 26/03/1997 at the European Commission-DG XIII D/1 (Telecommunications, information market and exploitation of research).

These two issues are dealt with a new semantic distance that is useful in the case of continuous, convex membership functions also allowing a definite integration.

The whole NAIADE procedure can be divided into four main steps:

1. pair wise comparison of alternatives according to each criterion,
2. aggregation of all criteria,
3. ranking of alternatives,
4. social conflict analysis.

From the mathematical point of view, an interesting characteristic of NAIADE is the possibility to fully control the degree of compensability allowed in the aggregation procedure.

The idea behind social conflict analysis is that criteria and criterion scores are not determined directly by social actors. The impact matrix is a result of a technical translation operationalized by the technical team. Even if the criteria are exactly the ones agreed with the social actors, the determination of the criterion scores is a technical issue independent of their preferences. This is one of the main reasons why it is desirable to combine a *social impact matrix* with the usual technical impact matrix normally used in a multi-criterion exercise.

Two issues are connected with all the outranking methods, as well as with other approaches based on pair-wise comparisons. First, Arrow's axiom of independence of irrelevant alternatives is not respected; thus the phenomenon of rank reversal may appear (i.e. the preference between ***a*** and ***b*** can change in function of the fact that a third option ***c*** is considered or not)[3]. Second, the Condorcet paradox may appear, i.e. alternative ***a*** may be ranked better than ***b***, ***b*** better than ***c*** and ***c*** better than ***a***. In addition, there is a problem specifically connected with the outranking approach. That is the necessity to establish a large number of "preference parameters", i.e. indifference and preference thresholds, concordance and discordance thresholds and weights. This may cause a loss of transparency and consistency in the model. In the framework of SMCE, outranking approaches are an interesting assessment framework, but to guarantee consistency with the social process behind the problem structuring, the mathematical aggregation rules need to be kept as simple as possible (see Munda, 2008 for a deeper technical discussion on this issue).

[3] Arrow's axiom of "the independence of irrelevant alternatives" states that the choice made in a given set of alternatives *A* depends only on the ordering made with respect to the alternatives in that set. Alternatives outside *A* (irrelevant since the choice must be made within *A*) should not affect the choice inside *A*. Empirical experience does not generally support this axiom. The issue of the independence of irrelevant alternatives is particularly important and tricky when pair-wise comparisons are used. To clarify this point, let's imagine a football championship. To determine the winner all the teams have to compete pair-wise. Then we need to know the performance of each team with respect to all the others, e.g., how many times a given team won, lost or was even. By using this information, we can finally determine who won the championship. Let's now imagine that when the championship is about to end and the team ***X*** is going to win (e.g. Barcelona), a new team ***Y*** is created (e.g. in Madrid). Would it be acceptable to allow this new team ***Y*** to play directly with ***X***? Would the supporters of team ***X*** accept that if ***Y*** wins, then ***Y*** will also win the championship? Of course not! This example seems to give a clear answer to our problem, but let's now imagine that instead of ranking football teams, our problem is to evaluate the performance of universities. Let's imagine that a study is almost finalized, and university ***A*** is going to be top ranked; however the study team discovers that an important university institution ***Z*** was not present in the original data set. Now the question is: can we just compare ***A*** with ***Z*** or do we have to make all the pairwise comparisons again? Now the answer is less clear cut. Moreover, let's imagine that the ranking at time *T* (without ***Z***) ranks university ***A*** better than ***B*** and that at time *T+1* (when ***Z*** is considered in the pair-wise comparisons) ***B*** is ranked better than ***A*** just because ***Z*** is taken into consideration! Can this result be acceptable? To answer this question in a definitive manner is very controversial. What we can say for sure is that if pair-wise comparisons are used, it has to be accepted the assumption that the irrelevant alternative ***Z*** (irrelevant for the evaluation between ***A*** and ***B***) can indeed change the relative evaluation of ***A*** and ***B***. This phenomenon is called "rank reversal".

4 An Illustrative Example

Let us consider as an illustrative example, an IA on modernising VAT for cross-border B2C e-Commerce developed recently. The impacts of the various options considered are summarised in the following impact matrix, showed in Table 11.

Key impacts	Option1	Option2	Option3	Option4	Option5	Option6
A – Economic and Competitiveness impact						
Impact on Member States						
Member States' revenues from intra-EU trade	=	-	+	+++	+++	+++
Cost for Member State to implement	=	-	-	- -	- -	- -
Effects on the volume and value of imports from third countries	=	--	+	+++	+++	+++
Impact on businesses						
Administrative burden	=	- -	-	++	+++	+++
Competition and growth in the EU						
Effects on intra-EU e-Commerce for goods and services	=	- - -	-	+	++	++
Effects on intra-EU e-Commerce prices	=	-	-	+	++	++
Effects on intra-EU e-Commerce value	=	+	++	++	++	++
Compliance						
Effects on Compliance	=	- - -	- - -	+ +	+ + +	+ +

B – Effectiveness of Options vs Policy Objectives						
Minimising burdens attached to cross-border e-Commerce arising from different VAT regimes.	=	- - -	- -	+ +	+ + +	+ + +
Providing a level playing field for EU businesses.	=	+	+	+ + +	+ + +	+ + +
Facilitating the monitoring of compliance and the fight against fraud for Member States' authorities.	=	- -	- -	+ +	+ + +	+ + +
Ensuring that VAT revenues accrue to the Member State of the consumer	=	+	+	+ + +	+ + +	+ + +
C – Coherence of options vis-à-vis the DSM Strategy						
Extending the current MOSS to intra-EU and 3rd country online sales of tangible goods	=	- - -	- - -	+ + +	+ + +	+ + +
Introducing a common EU-wide simplification measure (VAT threshold) to help small start-up e-commerce businesses.	=	- - -	+ +	+ + +	+ + +	+ + +
Removing the VAT exemption for the importation of small consignments from suppliers in third countries.	=	+	+	+ + +	+ + +	+ + +
Allowing for home country controls including a single audit of cross-border businesses for VAT purposes.	=	- - -	- - -	+	+ + +	+ +

D – Key indicators						
VAT Revenues (EUR)	137 bn	136.95 bn (- 0.05bn)	137.45 bn (+0.35bn)	144 bn (+7bn)	144 bn (+7bn)	144 bn (+7bn)
Business Compliance		4.7 bn	4.6 bn	2.4 bn	1.9 bn	2.1 bn
Costs (EUR)	4.2[78] bn	(increase of 0.5bn)	(increase of 0.4bn)	(decrease of 1.8 bn)	(decrease of 2.3 bn)	(decrease of 2.1bn)
Business Compliance Costs (%)	=	**+ 12%**	**+ 7%**	**- 43%**	**- 55%**	**- 51%**
Overall assessment	Does not meet objectives	Does not meet objectives	Does not meet objectives.	**Partially meets objectives**	**Fully meets objectives**	**Fully meets objectives**

Legend
+++ much better suited ++ better suited + slightly better suited
= no difference
- less suited - - slightly less suited - - - much less suited

Table 11. Impact Matrix of the Illustrative example

Source: SWD(2016) 379 final, pp. 46-48

In the original study, this impact matrix is commented on qualitative grounds[4] (without using any formal aggregation procedure) and it is concluded that *"**option 5** is considered to be the most positive as a business established in a Member State can make supplies to a customer in another Member State under broadly the same rules as a domestic transaction, the VAT rate applicable being the only exception. This option reduces overall compliance costs for business by 55% and evidence points to this option between the optimum one in terms of meeting the overall general and specific objectives of the proposal"* (p. SWD(2016) 379 final, 48).

Let's now see if this conclusion is corroborated by using a mathematical aggregation rule. By applying NAIADE, the ranking shown in Figure 2 is obtained. Each option is characterised by its strength, i.e. credibility that a number of options is worse than the one considered (positive flow Φ+) and weakness i.e. credibility that a number of options is better than the one considered (negative flow Φ-). The intersection between these two evaluations is providing the final ranking. When two options are not connected by an arrow, the situation described is a so-called *incomparability relation*, i.e. according to the information available, no clear relation of preference or indifference between these two options can be derived. Overall, in this case it seems that it is possible to state that the ranking is very clear: option ***5*** is the best choice followed by ***6*** and ***4;*** the set of options ***1, 2*** and ***3*** is clearly the worst one. However, by looking at the degrees of strength and weakness, one can realise that the difference among options ***4, 5*** and ***6*** is not very significant. More information can be obtained by checking the pairwise comparisons, which allow one to be fully aware of the mutual weaknesses and strengths on each single evaluation criterion.

[4] One should note that this qualitative approach is more a rule than an exception. In 17 IAs examined (SWD(2016) 392 final, SWD(2016) 152 final, SWD(2017) 31 final, SWD(2016) 211 final, SWD(2016) 434 final, SWD(2016) 410 final, SWD(2016) 418 final, SWD(2016) 173 final, SWD(2016) 303 final, SWD(2017) 26 final, SWD(2016) 315 final, SWD(2016) 193 final, SWD(2016) 468 final, SWD(2016) 17 final, SWD(2015) 135 final, SWD(2017) 202 final, SWD(2016) 379 final), only one (SWD(2016) 211 final) uses quantification (in the form of a mathematical aggregation rule) in ranking the considered options. All the other ones are based on a qualitative analysis of the various impacts, like the one described here.

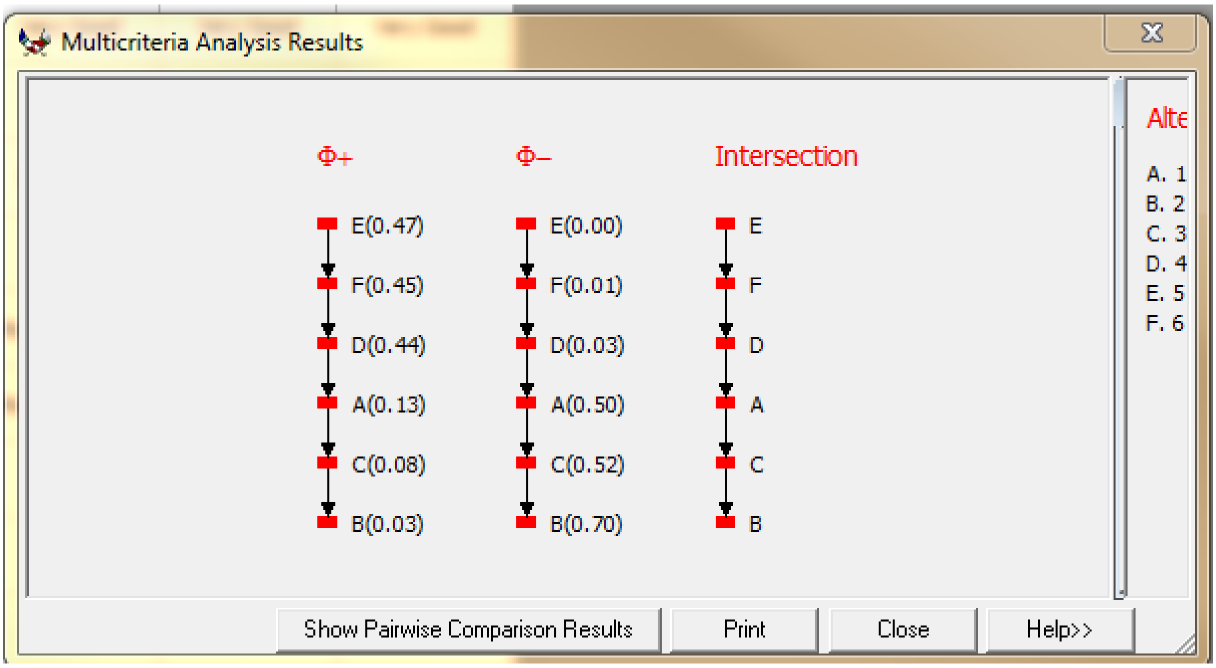

Figure 2. Multi-criteria ranking of all options shown in Table 11

In our example, the pairwise comparisons supply the results contained in Figures from 3 to 5 and all the ones contained in Annex 1. The first two columns report results on the credibility of the statement that the overall evaluation considers an option better, equal or worse than another one. In general an assessment is considered "credible" if its "degree of truth" is higher than 0.5. The third column provides the credibility of the evaluation referred to each single evaluation criterion. In this way, e.g. by looking at Figure 3, it is possible to deduce that options ***4*** and ***5*** are very similar (in technical terms, they are i*ndifferent*); this statement is corroborated by the first two columns. If one looks at the performance on each of the single criteria, it is possible to immediately see that only *criterion 19* (business compliance costs) is in favour of option ***5***, while all the other criteria evaluate options ***4*** and ***5*** as equal. By looking at Figures 4 and 5, we can indeed deduce the same type of evaluation: options ***4*** and ***6*** and ***5*** and ***6*** are basically equal. Between options ***4*** and ***6*** all criteria are in favour of an indifference relation, with the exception of criterion 19, which shows a clear preference of option ***6*** over ***4***. Between options ***5*** and ***6*** all criteria are in favour of an indifference relation, with a very weak credibility of a marginal preference of ***5*** over ***6*** again on criterion 19. Overall, on the grounds of this analysis, we can safely state that all three options are very similar with a slight preference attached on option ***5*** by criterion 19 only. This means that concluding that the best option is ***5*** implies to attach a very high weight to the criterion "business compliance costs" *implicitly*. By looking at the pairwise comparison shown in Annex 1, one can easily see that when comparing options ***4, 5*** and ***6*** with the other three, their preference is very clear, thus the clear evaluation seems to be that the set of options ***4, 5*** and ***6*** is definitely to be preferred to the set composed by options ***1, 2*** and ***3***.

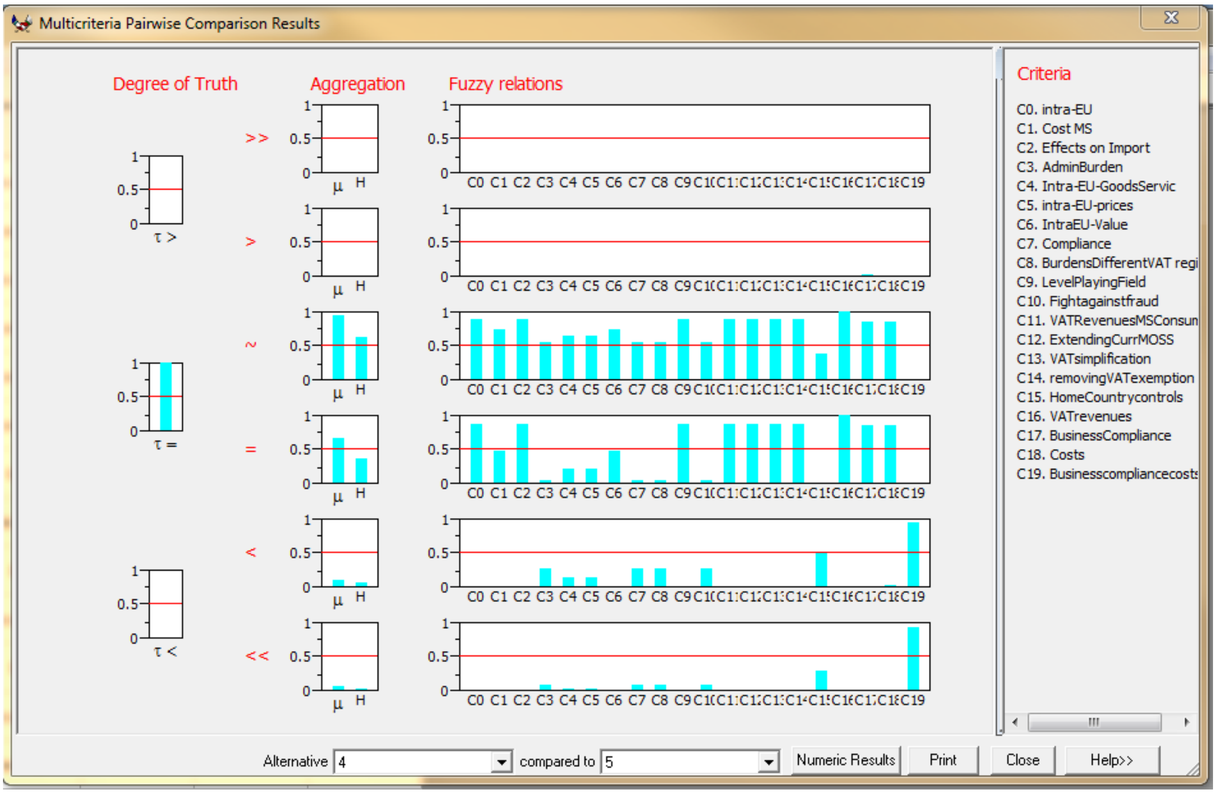

Figure 3. Pairwise comparison between options 4 and 5

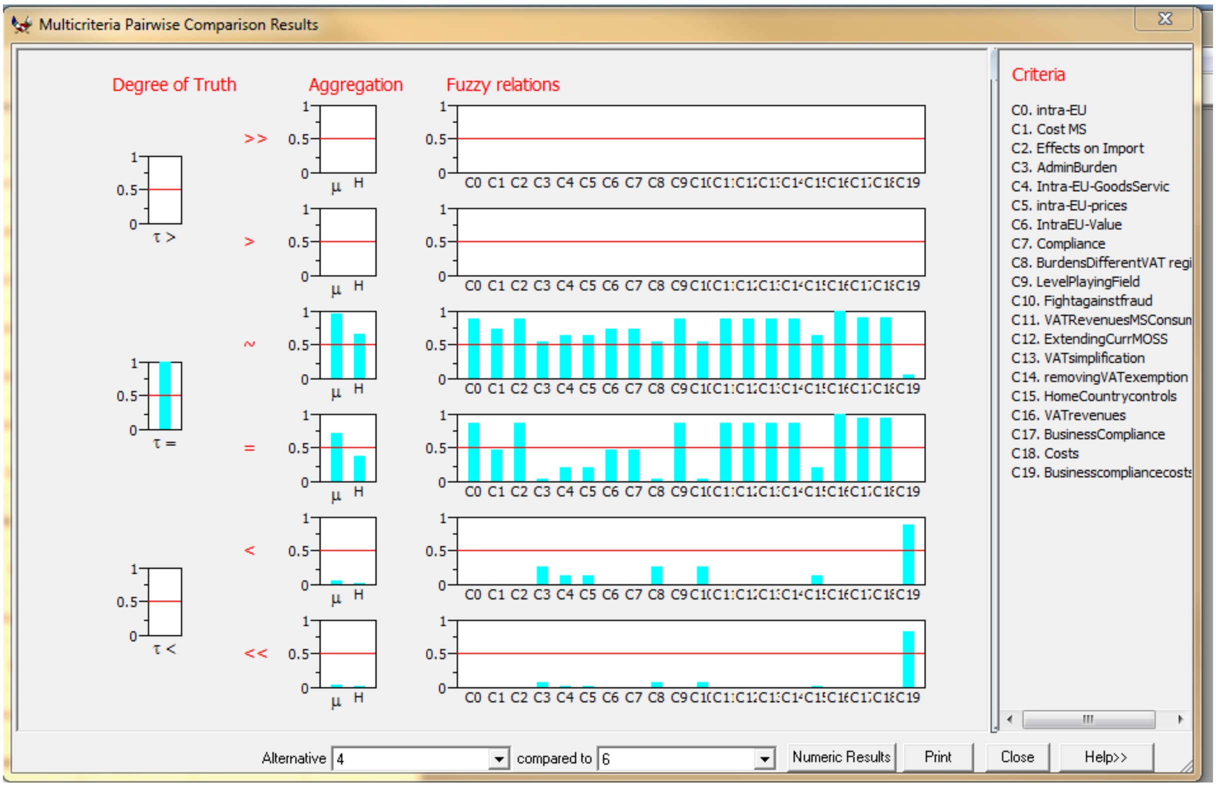

Figure 4. Pairwise comparison between options 4 and 6

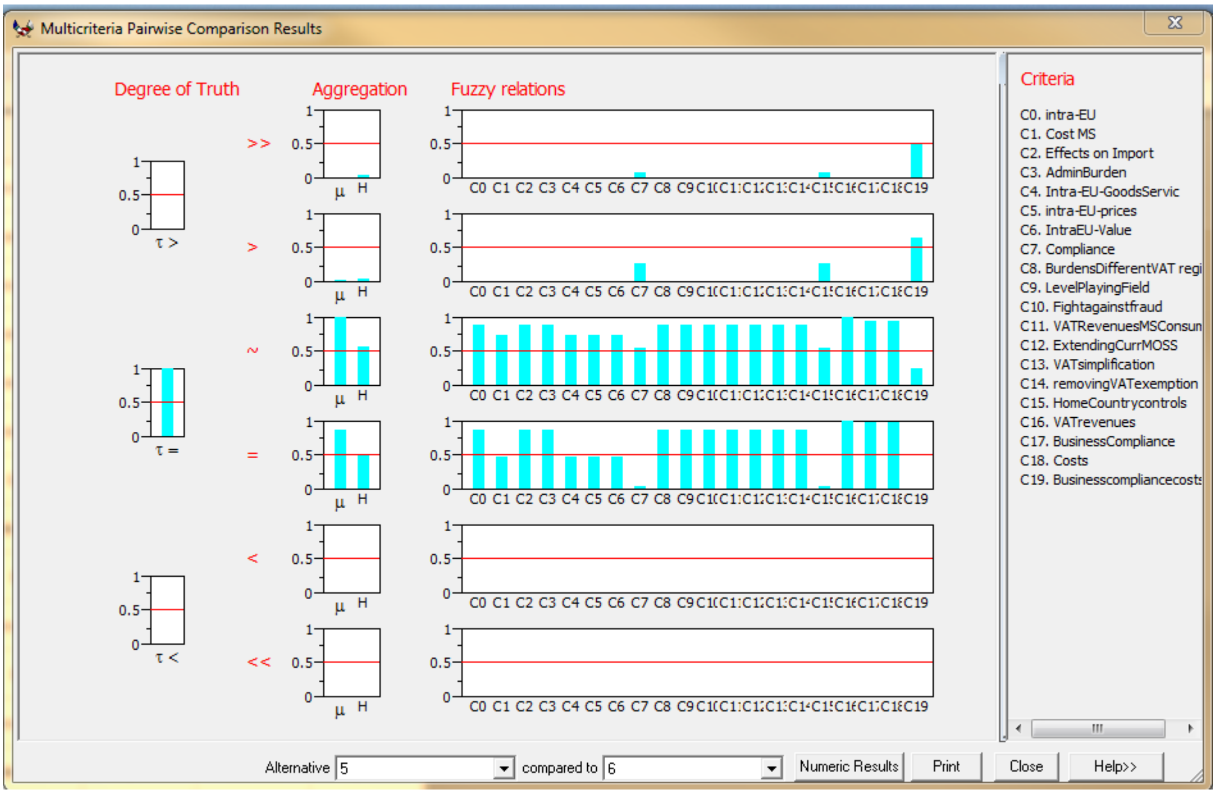

Figure 5. Pairwise comparison between options 5 and 6

To further clarify the preference structure, it is advisible to perform a robustness analysis according to criteria subsets (from A to D) illustrated in the original IA study.

A) *Economic Impacts*

According to this criteria subset, options ***4, 5*** and ***6*** present an incomparability relation among them, but it is clear that the three of them perform better than options ***1, 2*** and ***3***.

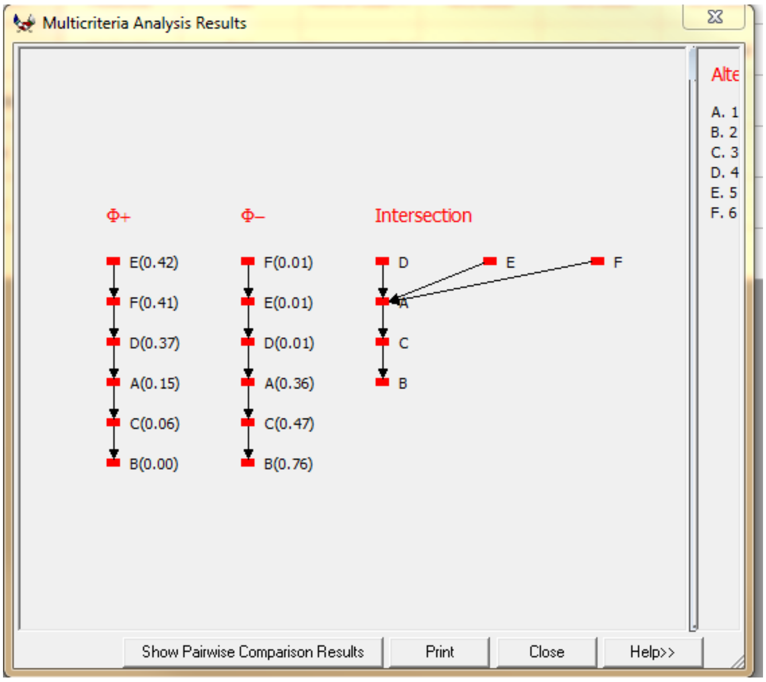

Figure 6. Multi-criteria ranking of all options according to the criteria subset A

B) Effectiveness

According to the criteria subset B, options ***4, 5*** and ***6*** again present an incomparability relation among them and the three of them perform better than options ***1, 2*** and ***3***.

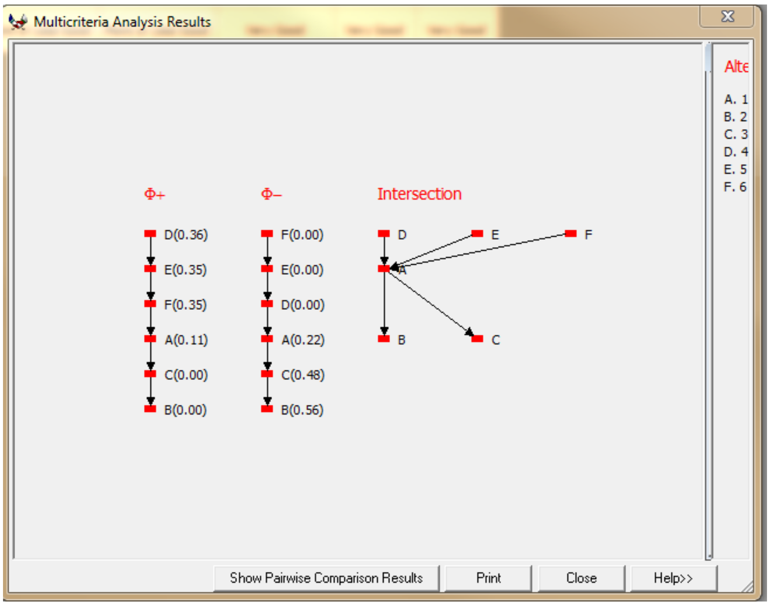

Figure 7. Multi-criteria ranking of all options according to the criteria subset B

C) Coherence

According to this criteria subset, options ***5*** is better than option ***4***, while option ***6*** presents an incomparability relation with both of them, the statement that they all perform better than options ***1, 2*** and ***3*** is corroborated.

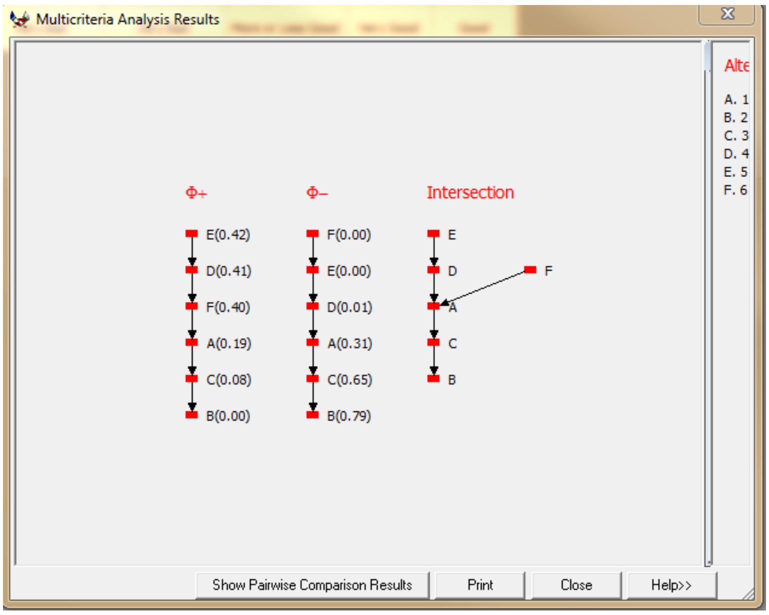

Figure 8. Multi-criteria ranking of all options according to the criteria subset C

D) Key Indicators

In this case, the ranking becomes clearer, a preference towards option ***5*** is derived. It becomes evident that this is the set of criteria that creates the preference ralation in favour of ***5***.

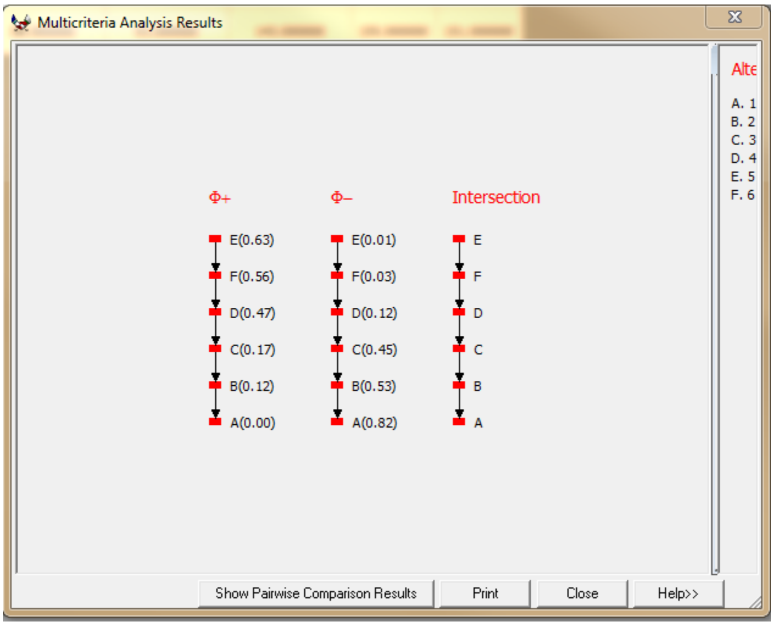

Figure 9. Multi-criteria ranking of all options according to the criteria subset D

As one can see, overall the statement that option ***5*** is a reasonable choice has been corroborated by the multi-criteria analysis, but it is also been clarified that this preference towards ***5*** is not very strong. All arguments to explicitate the preference structure to solve the impact matrix have been made clear, and no doubt the degree of transparency becomes much higher. Of course, the preference of option ***5*** over the other ones can also be justified by using other arguments than the ones contained in Table 11, but this possible additional use of criteria has to be made tansparent.

5 Conclusions

There is no doubt that *ex-ante* Impact Assessment is multidimensional in nature, and as a consequence social multi-criteria evaluation (SMCE), which has been explicitly designed for public policy, can be a very useful methodological and operational framework.

By using well established concepts in the decision theory literature, the option comparison step of Commission IAs, can be more homogeneous, transparent and its hierarchical structure be made clearer.

The use of mathematical aggregation rules in *ex-ante* IA studies, has al least three important justifications:

1. As clearly proved by Arrow's theorem, even when using mathematical rules, the ranking of policy options cannot be straightforward. *A fortiori,* the risk of deriving somewhat wrong rankings is much higher when qualitative reasoning only is used.
2. Even when the qualitative reasoning leads to correct conclusions, like the illustrative example we have examined, the use of mathematical aggregation rules brings more information and thus more *transparency* into the analysis.
3. When using mathematical rules, *consistency* between the problem structuring and the ranking of policy options is guaranteed, this makes the overall IA study much more defensible.

References

Arrow, K.J. (1963) Social choice and individual values. 2d edition, Wiley, New York.

Arrow, K.J., Raynaud, H. (1986) Social choice and multicriterion decision making. M.I.T. Press, Cambridge.

Brans, J.P., Mareschal, B. and Vincke, Ph. (1986) How to select and how to rank projects. The PROMETHEE method. European Journal of Operational Research, vol.24, pp. 228-238.

Figueira, J., Greco, S. and Ehrgott, M. (eds.) (2016) Multiple-criteria decision analysis. State of the art surveys. Springer International Series in Operations Research and Management Science, New York.

Moulin, H. (1988) *Axioms of co-operative decision making.* Econometric Society Monographs, Cambridge University Press, Cambridge.

Munda, G. (1995) Multicriteria evaluation in a fuzzy environment. Physica-Verlag, Contributions to Economics Series, Heidelberg.

Munda G. (2004) "Social multi-criteria evaluation (SMCE)": methodological foundations and operational consequences, European Journal of Operational Research Vol. 158, Issue 3: 662- 677.

Munda G. (2008) Social multi-criteria evaluation for a sustainable economy, Springer, Heidelberg, New York.

Munda G. (2016) Beyond Welfare Economics: Some Methodological Issues, Journal of Economic Methodology, Volume 23, Issue 2, Pages 185-202.

Nijkamp, P., Rietveld, P. and Voogd, H. (1990) Multicriteria Evaluation in Physical Planning. Amsterdam: North-Holland.

Podinovskii V.V. (1994) Criteria importance theory, Mathematical Social Sciences, 27, pp. 237 - 252.

Roy, B. (1996) Multicriteria methodology for decision analysis. Kluwer, Dordrecht.

Roberts F. S. (1979) Measurement theory with applications to decision making, utility and the social sciences, Addison-Wesley, London.

Vansnick J. C. (1990) Measurement theory and decision aid- in Bana e Costa C.A. (ed.)- Readings in multiple criteria decision aid, Springer-Verlag, Berlin, pp. 81-100.

Vincke Ph. (1992) Multicriteria decision aid*,* Wiley, New York.

List of abbreviations and definitions

IA	Impact Assessment
MCE	Multi-Criteria Evaluation
SMCE	Social Multi-Criteria Evaluation

List of figures

List of tables

Annexes

Annex 1. Pairwise Comparisons of Policy Options

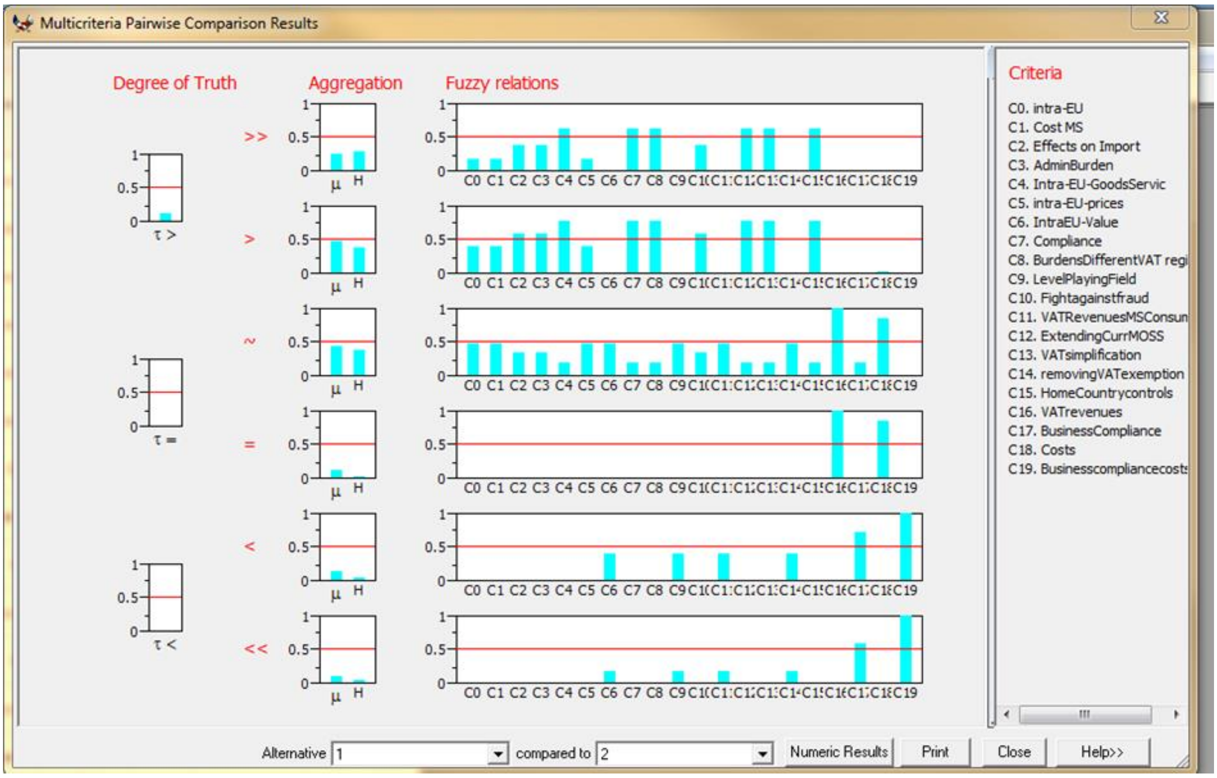

Figure A1. Pairwise comparison between options 1 and 2

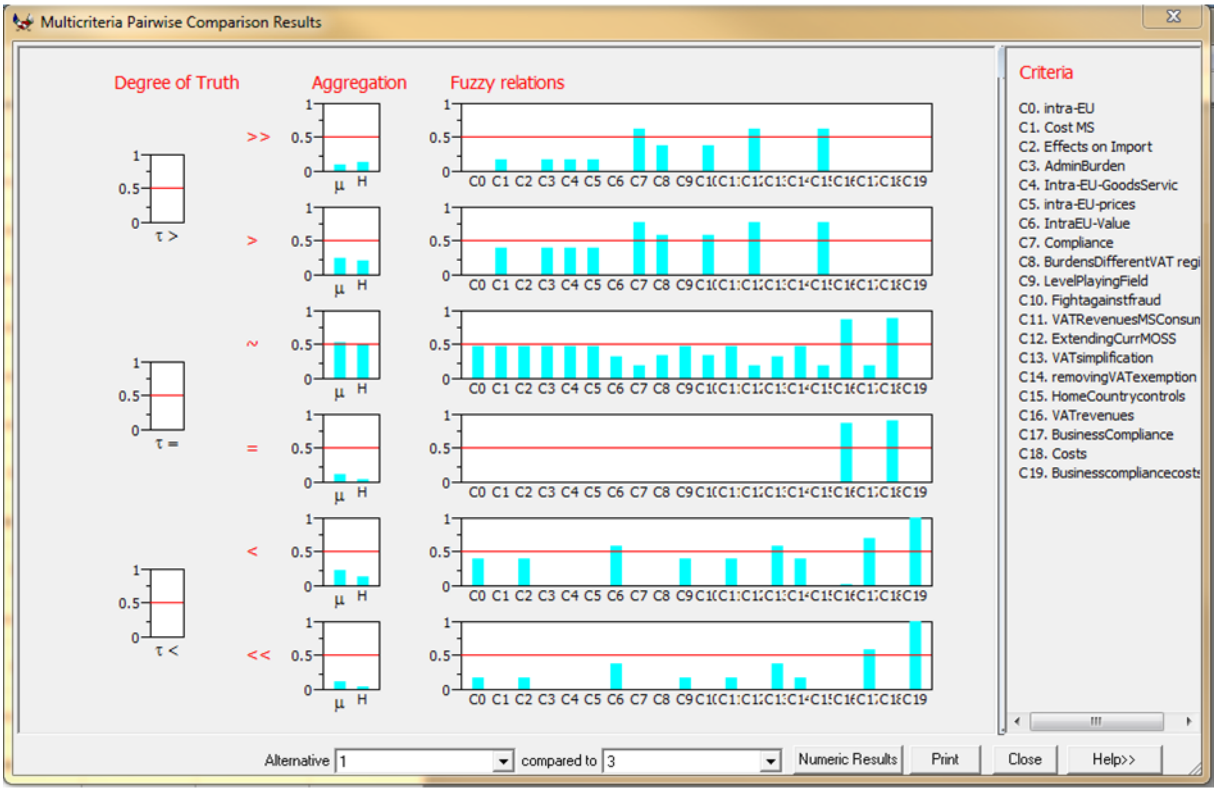

Figure A2. Pairwise comparison between options 1 and 3

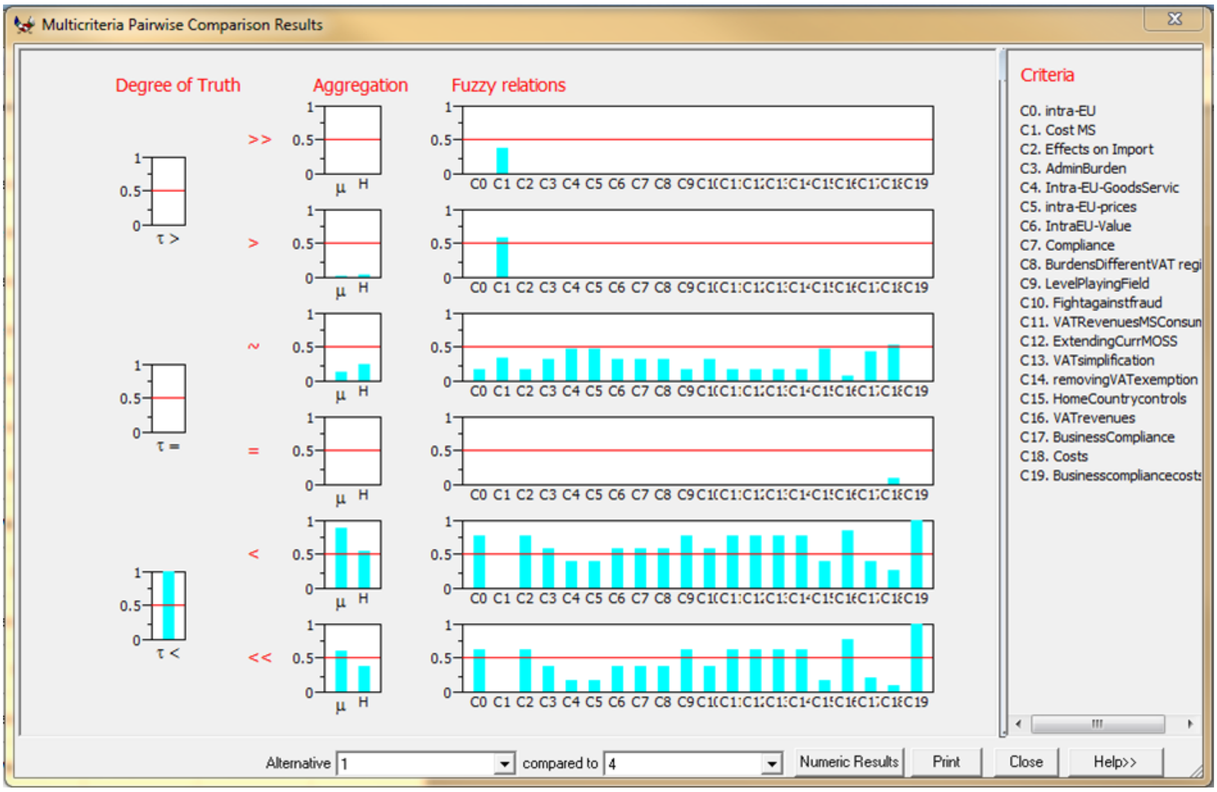

Figure A3. Pairwise comparison between options 1 and 4

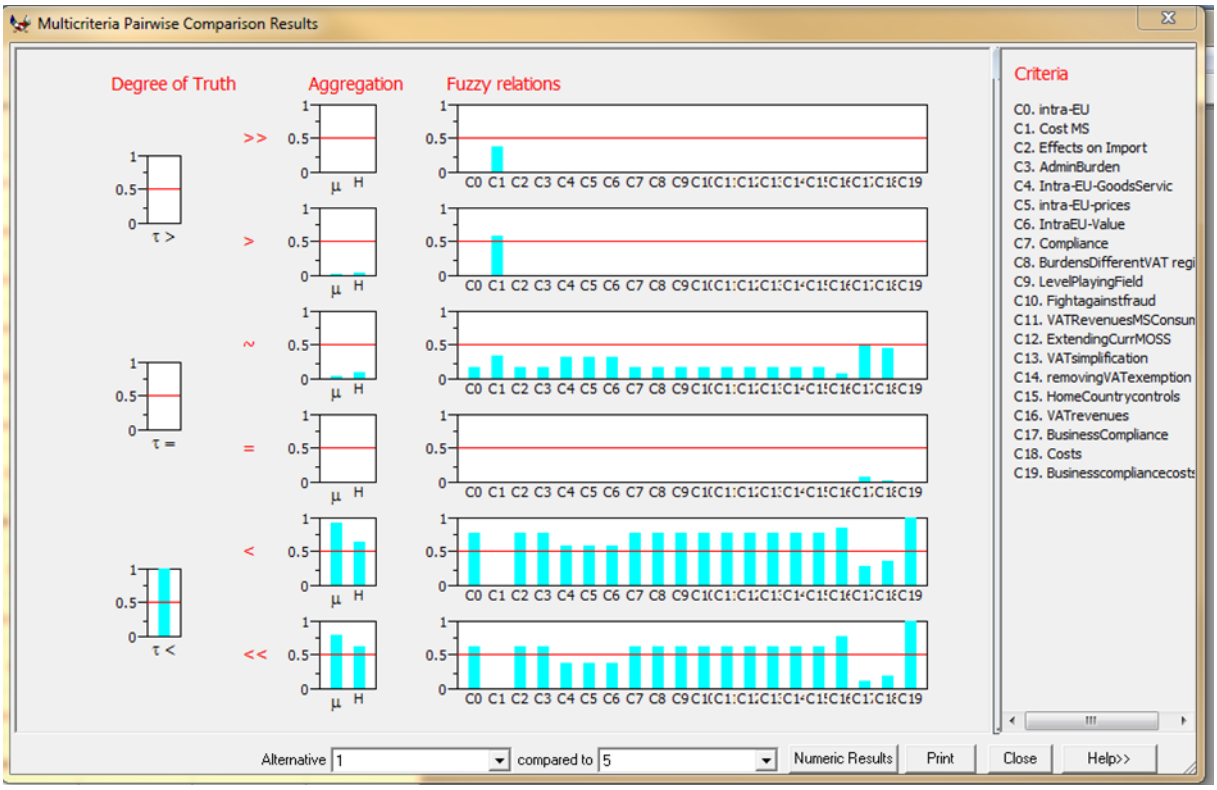

Figure A4. Pairwise comparison between options 1 and 5

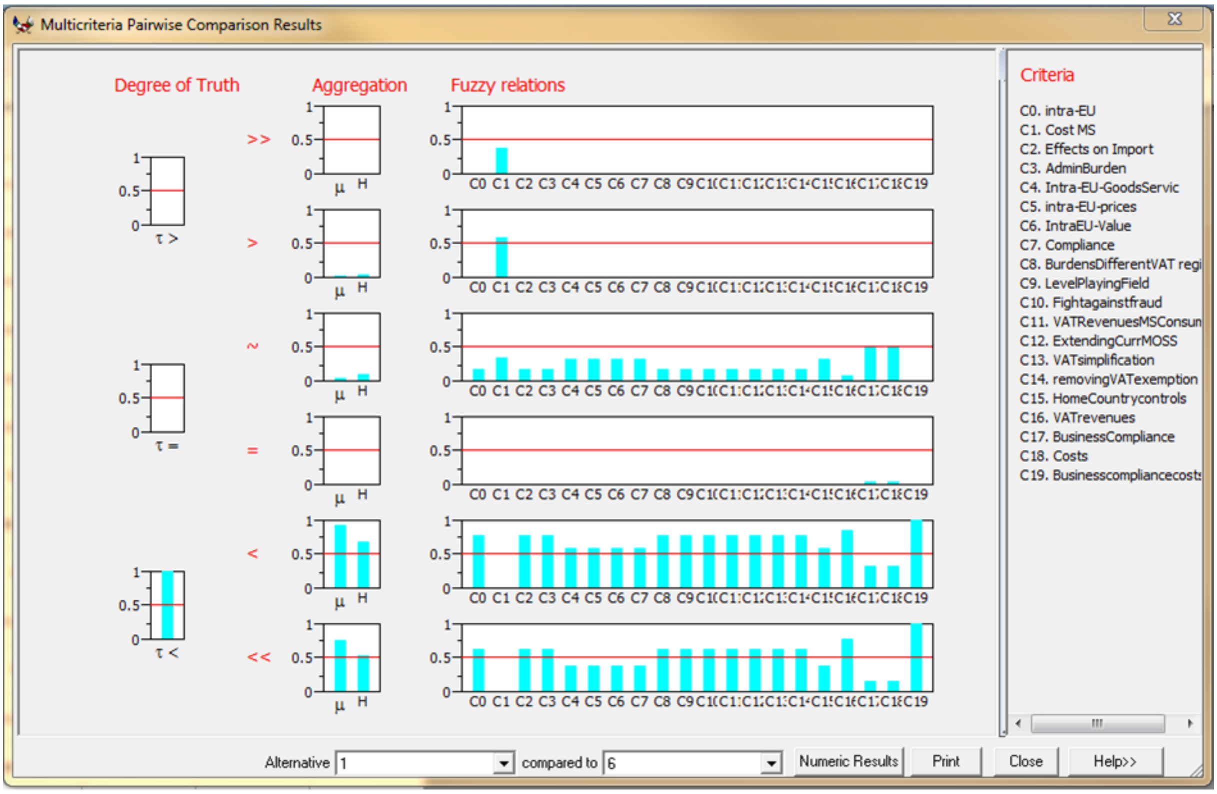

Figure A5. Pairwise comparison between options 1 and 6

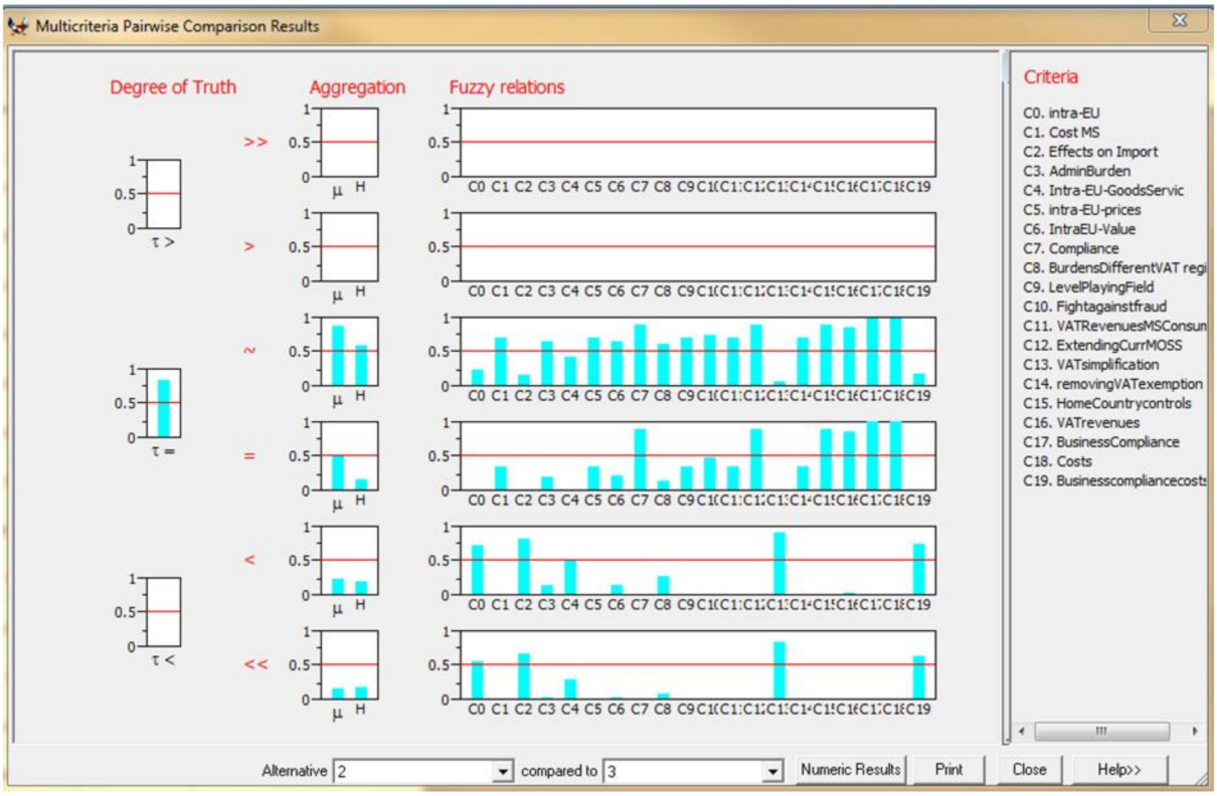

Figure A6. Pairwise comparison between options 2 and 3

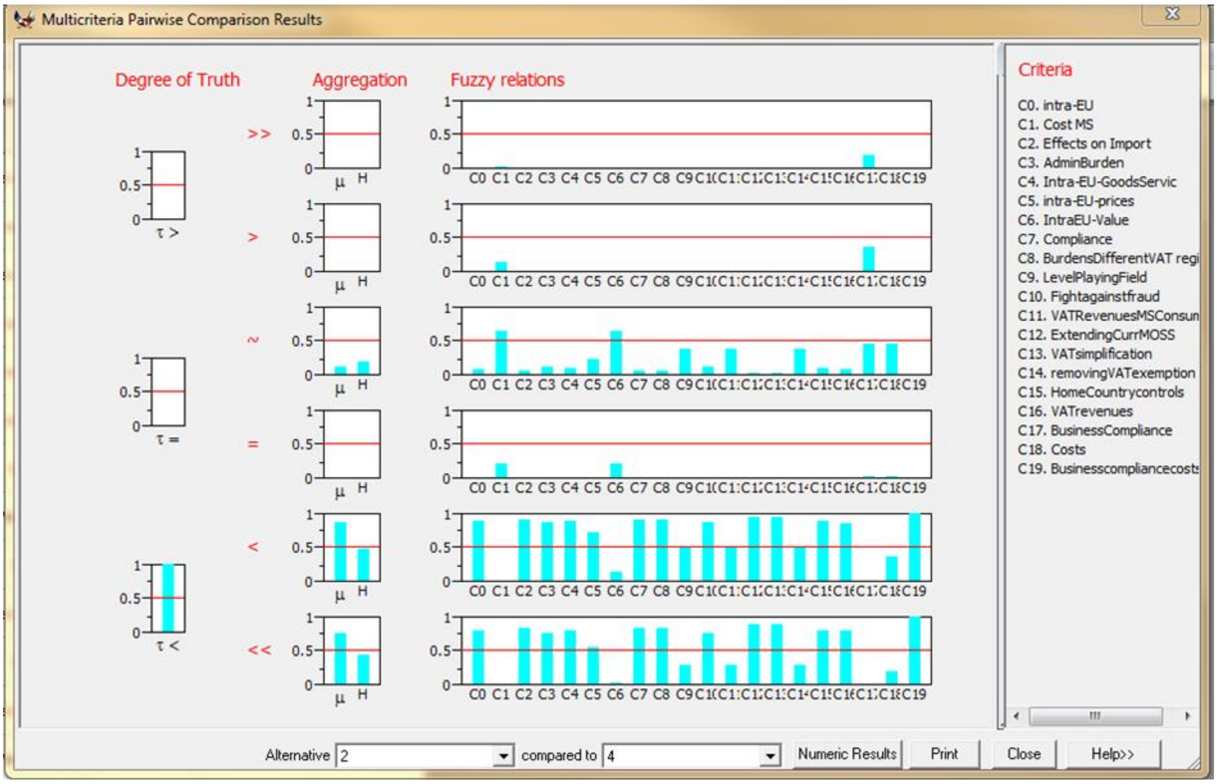

Figure A7. Pairwise comparison between options 2 and 4

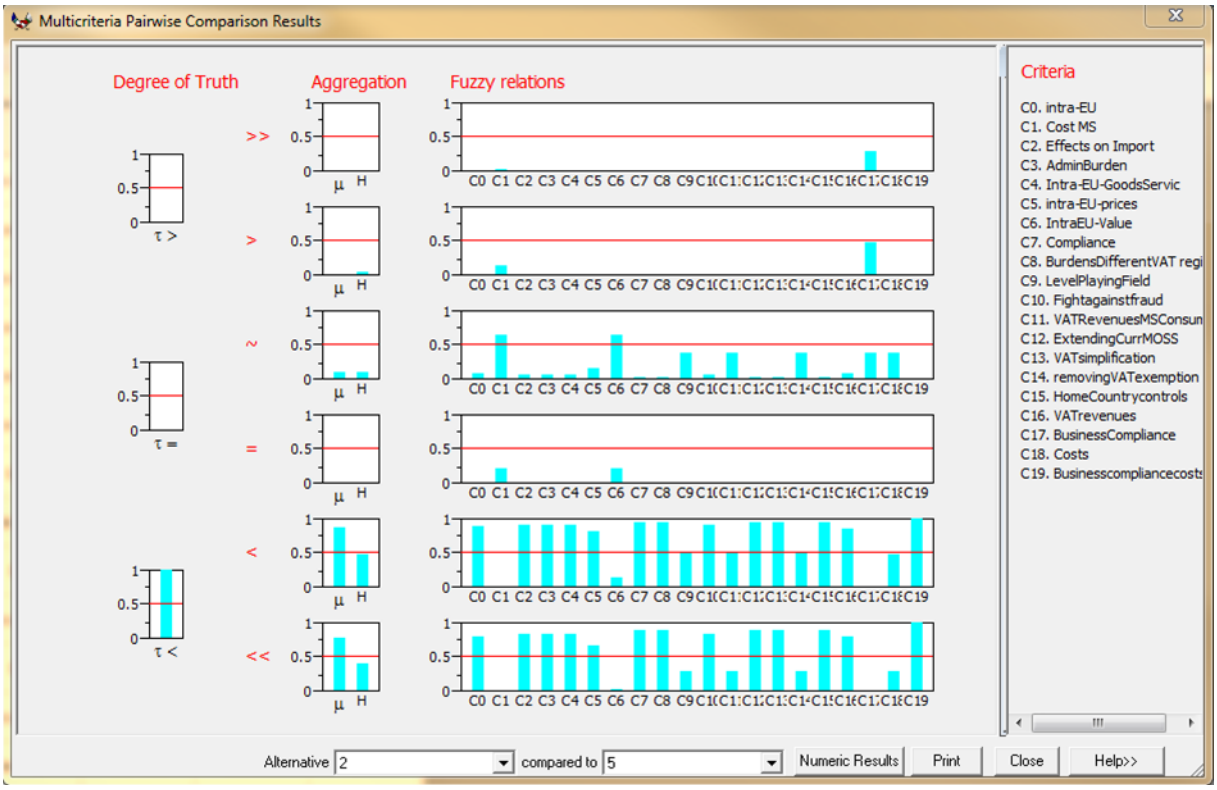

Figure A8. Pairwise comparison between options 2 and 5

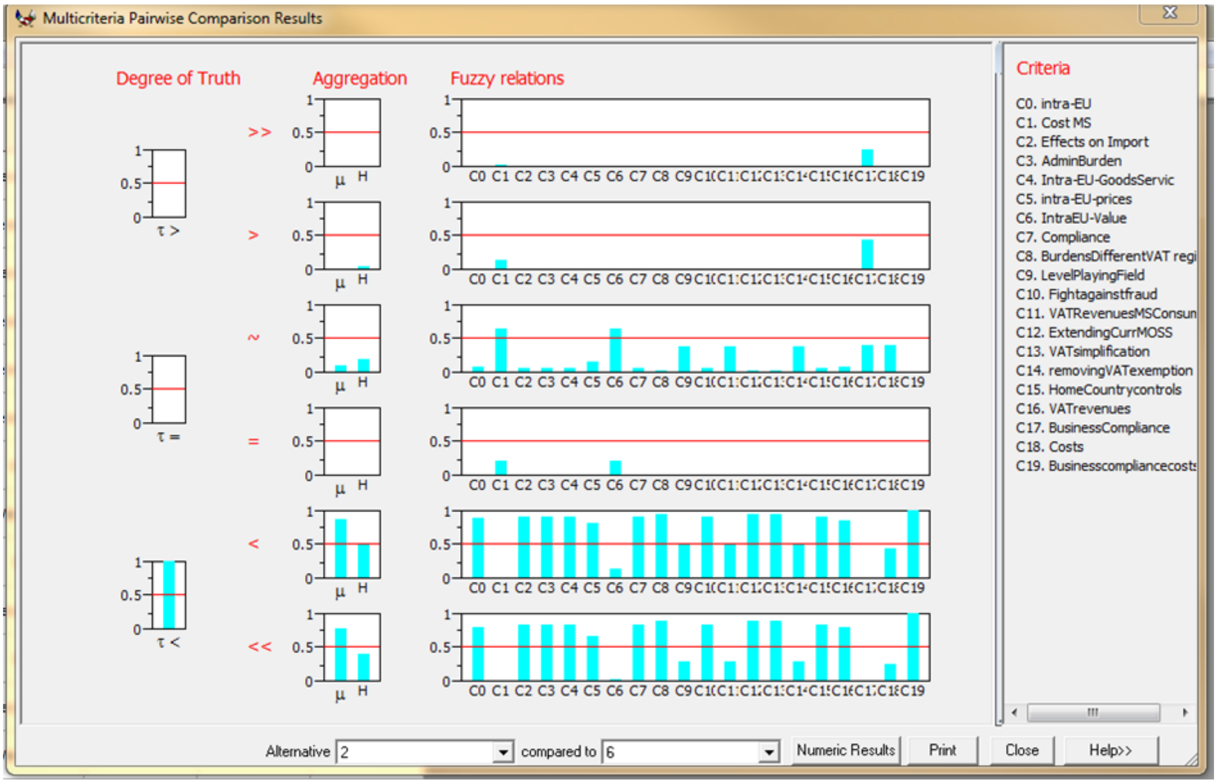

Figure A9. Pairwise comparison between options 2 and 6

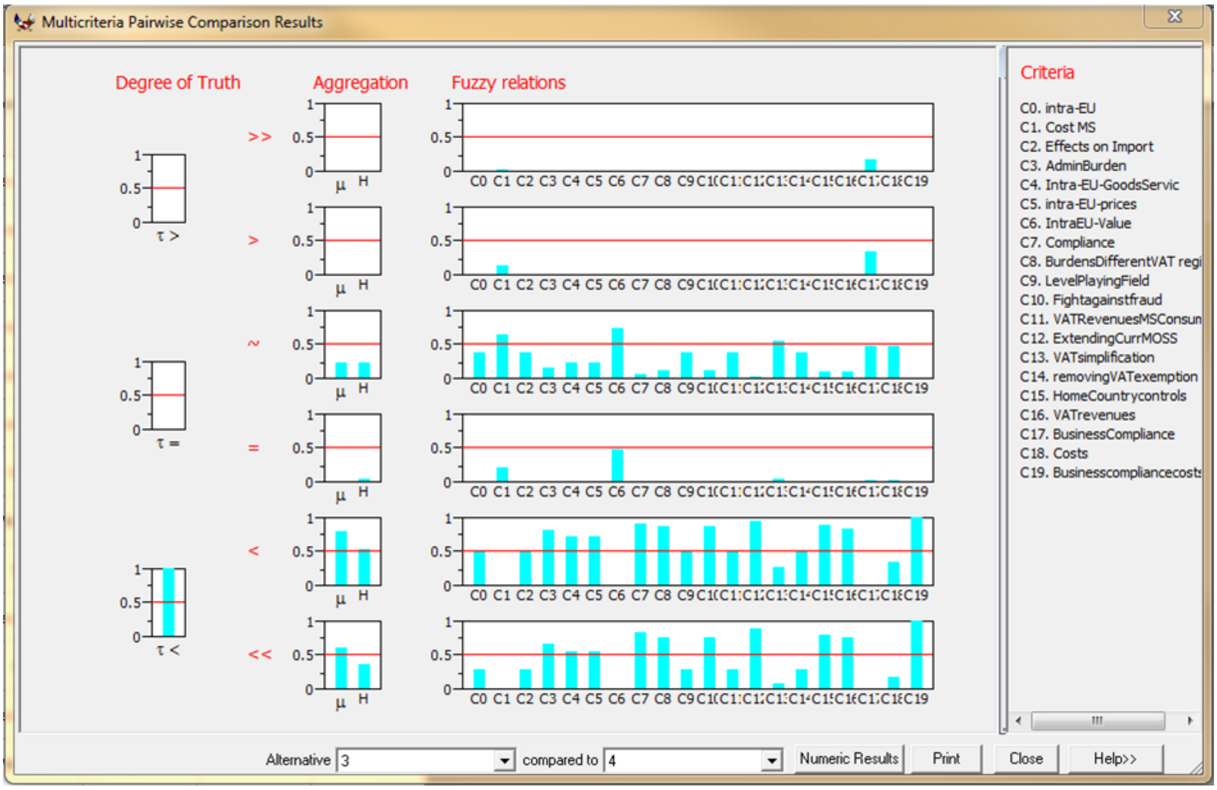

Figure A10. Pairwise comparison between options 3 and 4

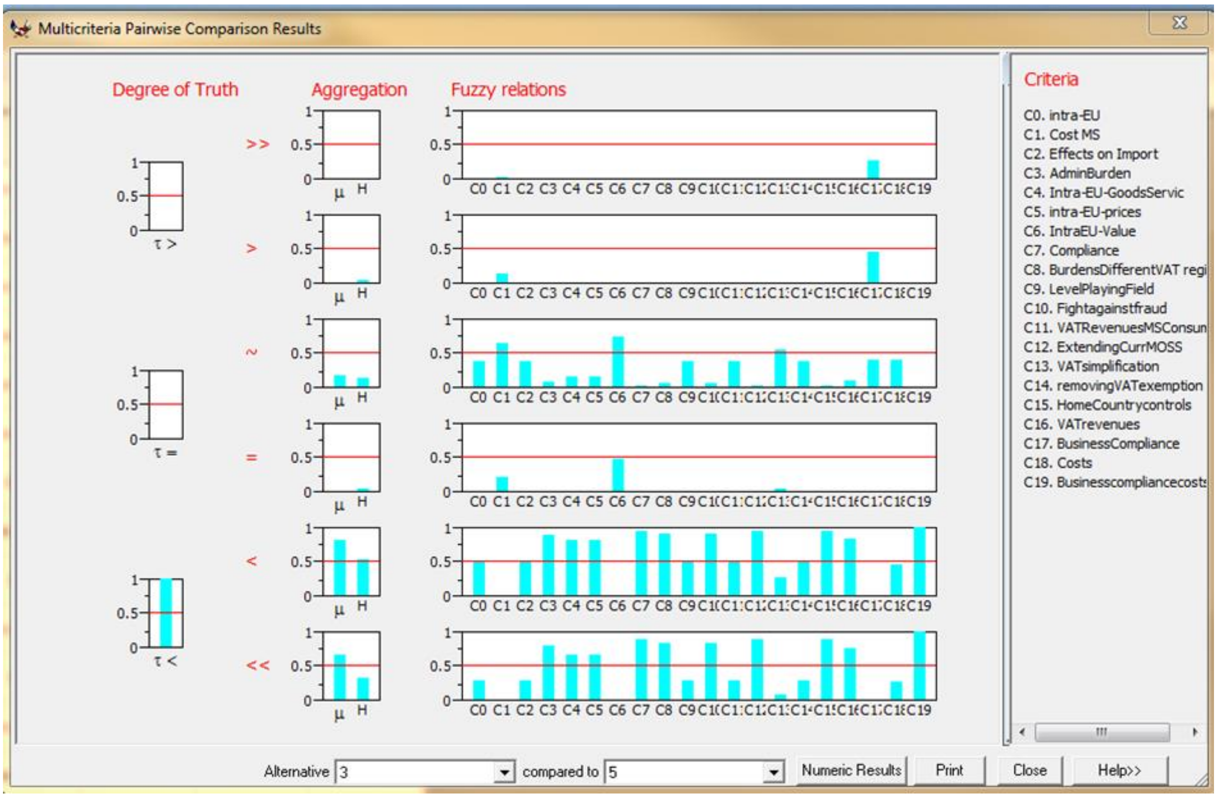

Figure A11. Pairwise comparison between options 3 and 5

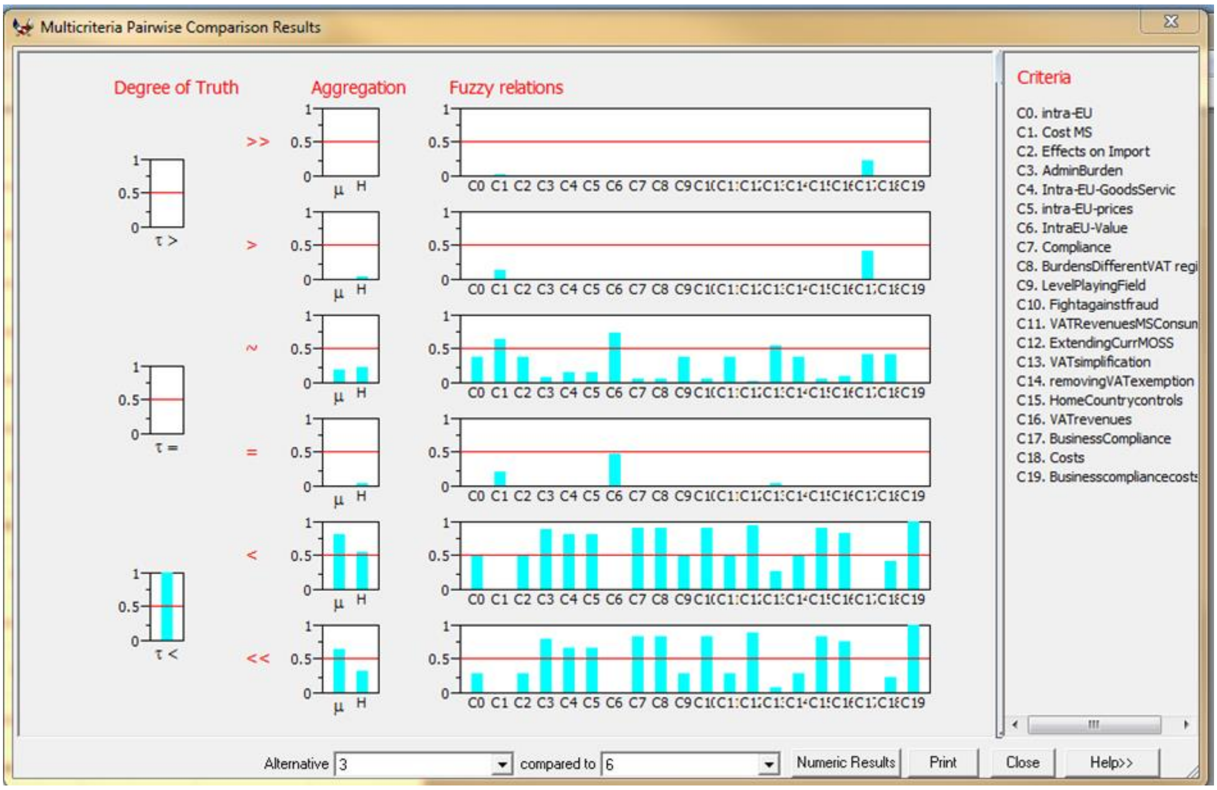

Figure A12. Pairwise comparison between options 3 and 6

Annex 2. Compensability and the Meaning of Criterion Weights

In the decision theory literature, the concept of weights as *symmetrical importance* is the following: *"… if we have two non-equal numbers to construct a vector in R^2, then it is preferable to place the greatest number in the position corresponding to the most important criterion."* (Podinovskii, 1994, p. 241; see also Munda, 2008 and Vincke, 1992).

This concept of weights as importance coefficients is very intuitive and it is how often weights are derived and used. However often there is a theoretical inconsistency in the way weights are actually used and their real theoretical meaning. In fact when one uses a compensatory approach in practice, such as the linear aggregation rule, one has to determine for each evaluation criterion, a mapping $\phi_i : x_i \rightarrow R$ which provides at least an interval scale of measurement and to assess scaling constants (i.e. weights) in order to specify how the compensability should be accomplished, given the scales ϕ_i between the different criteria (Roberts, 1979). Note that the scaling constants which appear in the compensatory approach depend on the scales ϕ_i, thus they do not characterise the intrinsic relative importance of each evaluation criterion.

There is unanimous agreement in the literature that the only method where weights are computed as scaling constants and there is no ambiguous interpretation is the so-called trade-off method starting with revealed preferences. No weight importance judgment is required in this method. The trade-off method can be briefly described as follows. Let's consider two options *A* and *B*, differing only for the criterion scores x_k and x_t. The problem is then to adjust the score x_k for *B*, in such a way that *A* and *B* become indifferent. Formally, it is:

$$I(A) = I(B) \Leftrightarrow I(x_1,...,x_k^{'},...,x_t^{'},...,x_n) = I(x_1,...,x_k^{''},...,x_t^{''},...,x_n) \Rightarrow \tag{A1}$$

$$\Rightarrow \sum_{\substack{i=1 \\ i \neq k,t}}^{N} w_i x_i + w_k x_k^{'} + w_t x_t^{'} = \sum_{\substack{i=1 \\ i \neq k,t}}^{N} w_i x_i + w_k x_k^{''} + w_t x_t^{''} \Rightarrow \tag{A2}$$

$$\Rightarrow w_k x_k^{'} + w_t x_t^{'} = w_k x_k^{''} + w_t x_t^{''} \tag{A3}$$

Equation (A3) is an equation in the unknown w_k and w_t. To compute the *N* weights as trade-offs, it is necessary to assess *N-1* equivalence relations which together with the usual normalisation constraint $w_i + \dots + w_n = 1$ determine a linear system of *N* equations in the *N* unknown weights. Of course if some uncertainty on the variable scores exists, this method cannot be applied.

As one can easily understand to assess weights as trade-offs, as it should be always done when e.g. using a linear aggregation rule, it is a much harder job than to use weights as importance coefficients. This is probably the main reason why the standard practice tends to use weights as importance coefficients, even in compensatory aggregation rules such as the linear one but unfortunately this practice is not defensible on theoretical grounds.

Vansnick (1990) showed that the two main approaches in multi-criteria aggregation procedures i.e., the compensatory and partial-compensatory ones can be directly derived from the seminal work of Borda and Condorcet. If one wants the weights to be interpreted as "importance coefficients" (or equivalently *symmetrical importance* of criteria) fully compensatory aggregation procedures must be avoided. From a social choice point of view, partial-compensatory rules are always Condorcet consistent rules.

KJ-NA-28752-EN-N

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office

doi:10.2760/909528
ISBN 978-92-79-72293-6